大数据分析技术及其在数字出版中的运用探究

摘　要：我国的经济社会不断发展，网络信息技术水平不断提升，其与社会各个行业的融合日益紧密。大数据分析技术是现代技术的代表，将其应用在出版行业中，可以收获良好的效果。为了发挥大数据分析技术的实用价值，必须推动新媒体出版行业的发展。本文将具体探讨大数据分析技术及其在数字出版中的运用，希望能为相关人士提供一些参考。

关键词：大数据分析技术；数字出版；运用

中图分类号：G23　　文献标识码：A

文章编号：1671-0134（2017）07-067-02　　DOI：10.19483/j.cnki.11-4653/n.2017.07.019

■文 / 马国宝

进入21世纪以来，我国的社会主义市场经济持续繁荣，现代技术发展更加迅速。在信息化时代中，数据信息是最具价值的资源。出版行业记录着大量的出版信息，为了实现行业发展，必须对数据信息进行分析，制定科学的发展决策。大数据分析技术改变了传统的新闻出版模式，催生了数字出版这一新模式。为了扩展数字出版的范围，必须探讨大数据分析技术的实践意义。

1. 大数据分析技术概述

1.1 内涵

所谓的大数据分析技术，就是以互联网技术作为依托，对数据进行储存、管理和精准分析，从中识别有用信息，自动剔除无用信息的现代化技术[1]。大数据分析技术具有突出的实用功能，得到了各个行业的普遍重视。

我国的经济社会不断发展，网络信息技术渗透了人们日常生活的方方面面。人们在生产生活中会接触到大量的数据信息，只有对海量数据信息进行分析，才能提升数据处理能力，实现自身的可持续发展。传统的信息处理手段具有滞后性，存在时间延缓的弊病，信息处理结果只能反映上一时段的情况。大数据技术弥补了传统信息处理方式的不足，将各种数据信息整合到了一起，可以对实时数据进行快速分析。

就内容来看，大数据技术包括对数据流进行预测、对数据库进行管理、对数据算法进行优化等。就方法来看，大数据技术包括数据收集方法、数据录入方法、数据处理方法等。就操作来看，大数据技术包括数据挖掘、数据库建构、数据统计等。

1.2 数据挖掘

数据挖掘是大数据分析技术中最重要的操作方法。数据挖掘又称数据开发，是对互联网中的海量数据进行开发，从中甄选出最具价值的实时信息。外部环境处在不断发展变化中，数据也具有极强的变化性，数据挖掘以数据实时特征作为基点，可以对数据资源进行类别化分析。具体来说，数据挖掘操作包括以下几个环节。

第一是划分数据的类型。在互联网中有大量的数据信息，为了方便管理，需要形成不同的数据库，记录不同类型的数据。就出版行业来看，数据信息包括用户信息、出版社信息、书籍信息等，数据发掘可以将数据分门别类地保存在数据库之中；第二是对数据进行细分。每一类数据都可以被划分为不同的层次，以用户信息为例，包含用户偏好信息、用户产品信息、用户属性信息等；以出版社信息为例，包含出版社营业信息、出版社配送信息、出版社销售信息等；以书籍信息为例，包含书籍作者信息、电子书籍信息、书籍促销信息等；第三是分析数据信息的特征。每类信息都有突出的个性化特征，数据挖掘可以将数据信息放在一起，从中提取出最具价值的特征信息；第四是对数据的回归分析。数据有一定的流动性，数据挖掘可以根据数据流方向，对出版行业的书籍销售趋势进行预测。

2. 数字出版概述

2.1 内涵

所谓的数字出版，就是依托网络数据信息的新型出版模式。数字出版以网络信息技术作为基础，可以弥补传统出版模式的不足之处。数字出版具有极强的系统性特征，不仅包括出版方式的数字化，还包括出版业务、用户消费的数字化等[2]。在采用数字出版的过程中，需要依靠互联网技术构建信息平台，在平台上为用户提供电子阅读书目。用户在阅读之前，需要先支付一定的阅读费用，与商家进行交易。

2.2 优势

首先，数字出版具有突出的便捷性特征。数字出版以网络信息技术作为支撑，储存了大量数据信息。在搜索数据的过程中，用户可以根据数据类型寻找到相应数据，节省大量搜索时间。出版行业的每个链条都有相关性，数字出版可以在网络平台上传递数据，实现信息的交互和共享。传统出版以人的操作为主，误差较大，耗费成本较高。数字出版以智

能操作为主，误差较小，耗费成本较低。从这个角度来看，数字出版已经成为出版行业发展的必然之势。

其次，数字出版具有突出的丰富性特征。就目前来看，我国的数字出版包括数字报刊、电子书、电子地图等多种产品，各种移动终端都可以加载上述产品。随着网络信息技术水平的不断提升，线下出版产品已经扩展到了线上。与传统出版相比，数字出版可以带给用户更加直观的服务体验，改变用户的娱乐方式。

再次，数字出版具有突出的发展性特征。数字出版以互联网技术作为依托，互联网技术不断发展，数字出版的发展速度也会不断加快。在信息化时代中，数字出版已经引起了大众的关注，成为最具活力的现代技术产物之一。大数据技术为数字出版注入了新鲜的血液，势必会推动当代出版行业的发展。

3. 大数据技术在数字出版中的运用

3.1 出版流程

大数据技术可以被应用在数字出版的各个环节之中，节省出版时间。具体来说，数字出版包括以下几个流程：第一是准备出版产品，对产品特性进行分析，分门别类进行管理；第二是制定产品营销策略，对市场情况进行分析，和大众媒体进行联合；第三是优化配置人力资源和物力资源，对每个工作环节的工作量进行测算，工作量越大，人力资源和物力资源的输出越多；第四是对用户进行集中管理，储存海量用户信息。比如，在准备出版产品的过程中，可以应用大数据技术中的 RDM 系统[3]。RDM 系统是智能网络管理平台，可以对出版产品进行集中管理，为了辅助 RDM 系统的工作，也可以构建 ERP 系统，并划分两个管理系统的工作内容。在系统产生数据后，要应用大数据技术，对实时数据进行精准分析。如果产品信息出现讹误，需要对其进行二度审核，并修正讹误信息。

数字出版流程有两条线索，第一条是业务流程线索，第二条是管理流程线索。出版流程不同，大数据技术的应用也呈现出差异。以业务流程为例，在设计业务流程时，需要利用大数据技术，考察市场消费情况和用户的个人偏好，并将各种信息记录在数据库中。为了提高业务流程的设计水平，可以形成网络信息平台，吸引更多设计师的注意。以管理流程为例，在设计管理流程时，需要形成数据分析部门，对商业数据进行搜寻和处理。同时要考察供需关系，对投入产品比重进行控制。

3.2 出版产品

大数据技术可以被应用在数字出版产品中。在物质生活日益丰富的今天，人们的精神需求日益旺盛，对出版产品提出了更高的要求。出版产品只有满足用户的多元需求，才能促进出版行业的可持续发展。大数据分析技术可以对用户需求进行调查，根据需求数据创新出版产品，为用户提供多种选择。在大数据技术的支持下，产品生产速度加快，产品的市场占有率不断上升，创造了更多的经济价值和社会价值。除了提供产品之外，数字出版还可以为用户提供线上服务。大数据技术重构了数字服务的内容，跨越了时间和空间的界限。用户只需要登录网络平台，就能和供应商实现交互。

为了实现经济效益的最大化，供应商必须采用有效的营销策略，而营销策略要以用户偏好作为基础。大数据技术可以为营销策略提供海量数据，帮助出版行业构建新的营销方式。一方面，大数据技术可以形成互动平台，让供应商和用户相对接；另一方面，大数据技术可以综合考量用户需要，为用户提供定制化的数字产品。

3.3 出版服务

大数据技术可以被应用在数字出版服务中。在大数据技术的支持下，数字出版更加智能，在很大程度上推动了我国出版行业和的发展。根据用户的历史数据，大数据技术可以为用户推送相关内容，避免用户在网络上进行重新搜索。用户的历史数据包括用户的消费偏向、用户的内容喜好、用户的个人评价等[4]。大数据技术需要按照主题、人物等，把产品划分为不同的类型，尽量和消费市场相贴合。

在提供服务的同时，还应该和用户建立好关系，让用户进行信息反馈。比如，在微信、微博等社交媒体中，应该汲取用户提出的意见和建议。大数据技术可以对数据进行智能分析，供应商应该利用数据挖掘功能，对用户体验进行评价，以此为依据调整自身的服务方式。

综上所述，我国的经济社会不断发展，出版行业进入快速发挥阶段。为了扩展数字出版的范围，必须探讨大数据分析技术的实践意义。媒

参考文献

[1] 王鹏，赵丽虹 . 大数据和新媒体技术推动城市规划转型初探 [A]. 中国科学技术协会、广东省人民政府 . 第十七届中国科协年会——分 16 大数据与城乡治理研讨会论文集 [C]. 中国科学技术协会、广东省人民政府，2015：7.

[2] 刘实然 . 面向出版行业的新媒体精准营销系统的设计与实现 [D]. 中国科学院大学（工程管理与信息技术学院），2015.

[3] 滕跃民，滕莉 . 新媒体的发展及技术与市场分析——“新媒体出版与传播高级研修班”综述 [J]. 出版与印刷，2016（04）：7-10.

[4] 杨淑娟，刘景景，沈阳 . 媒体微信公众平台服务发展现状及对策——基于“新媒体指数”大数据平台的分析 [J]. 新闻与写作，2015（02）：10-14.

（作者单位：河南科学技术出版社有限公司）